BIOLOGICAL Fib/T26 Suppl 1

FAO Fisheries Technical Paper No. 26 Suppl. 1
(Distribution restricted)

Fib/T26 Suppl. 1
MFS-Sampling

MANUAL OF SAMPLING AND STATISTICAL METHODS FOR FISHERIES BIOLOGY

COMPUTATIONS

by

K. P. Anderson

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 1965

DOCUMENTS OF THE BIOLOGY BRANCH OF FAO FISHERIES DIVISION

Documents which are not official FAO publications are issued in several series. They are given a restricted distribution and this fact should be indicated if they are cited. Most of them are prepared as working papers for meetings, or are summaries of information for use of member governments, organizations, and specialists concerned.

DOCUMENTS DE LA SOUS-DIVISION DE LA BIOLOGIE DE LA DIVISION DES PÊCHES DE LA FAO

Des documents qui ne figurent pas parmi les publications officielles de la FAO sont publiés dans diverses séries. Ils font seulement l'objet d'une distribution restreinte, aussi convient-il de le préciser lorsque ces documents sont cités. Il s'agit le plus souvent de documents de travail préparés pour des réunions, ou de résumés d'information à l'intention des gouvernements des pays membres, ainsi que des organisations et spécialistes intéressés. Ces séries sont les suivantes:

DOCUMENTOS DE LA SUBDIRECCION DE BIOLOGIA DE LA DIRECCION DE PESCA DE LA FAO

Esta Subdirección publica varias series de documentos que no pueden considerarse como publicaciones oficiales de la FAO. Todos ellos tienen distribución limitada, circunstancia que debe indicarse en el caso de ser citados. La mayoría de los títulos que figuran en dichas series son documentos de trabajo preparados para reuniones o resúmenes de información destinados a los estados miembros, organizaciones y especialistas interesados.

FAO Fisheries Report	**FIb/R (No.)**
FAO Fisheries Circular	**FIb/C (No.)**
FAO Fisheries Synopsis	**FIb/S (No.)**

Special groups of synopses are identified by symbols followed by classification numbers based on indexed code of " Current Bibliography ":

SAST	Data concerning certain species and fish stocks.
MAST	Information on methods and subjects.
OT	Oceanographic data.
IT	Limnological data.
CART	Information concerning fisheries and resources of certain countries and regions (FID/S).

Des catégories spéciales d'études synoptiques sont identifiées à l'aide de symboles suivis des chiffres de classification basés sur le code d'indexation de la « Current Bibliography »:

SAST	Données sur certaines espèces et populations de poissons.
MAST	Renseignements sur des méthodes et des sujets.
OT	Données océanographiques.
IT	Données limnologiques.
CART	Renseignements sur les pêcheries et les ressources de certains pays et régions (FID/S).

Grupos especiales de sinopsis se distinguen con las siglas siguientes, seguidas por números de clasificación que se basan en las claves de los índices de la « Current Bibliography »:

SAST	Datos relativos a ciertas especies y poblaciones.
MAST	Sinopsis sobre métodos y materias.
OT	Sinopsis sobre oceanografía.
IT	Sinopsis sobre limnología.
CART	Información sobre los recursos acuáticos vivos de algunos países y regiones (FID/S).

FAO Fisheries Technical Paper FIb/T (No.)

Special groups of Technical Papers are identified by:

RE	Indexed lists of experts and institutions drawn from Registers maintained by Biology Branch.
CB	Lists of Periodicals, special sections of " Current Bibliography for Aquatic Sciences and Fisheries," special bibliographies and papers concerning documentation problems.
MFS	Provisional editions of " FAO Manuals in Fisheries Science."

Some documents also have another identification, if, for example, they have been contributed to a meeting for which papers have been numbered according to another system.

Des catégories spéciales de documents techniques sont identifiées à l'aide des symboles suivants:

RE	Listes indexées d'experts et institutions tirées des registres tenus à jour par la Sous-division de la biologie.
CB	Listes de périodiques, des sections spéciales de la « Current Bibliography for Aquatic Sciences and Fisheries », des bibliographies particulières et des articles sur les problèmes de documentation.
MFS	Editions provisoires des « Manuels FAO de la science des pêches ».

Certains documents portent parfois d'autres numéros d'identification, par exemple s'ils ont été préparés pour une réunion dont les documents ont été numérotés selon un autre système.

Grupos especiales de documentos técnicos se identifican por las siglas siguientes:

RE	Listas índices de expertos y de instituciones tomadas de los registros que se llevan en la Subdirección de Biología.
CB	Listas de periódicos, secciones especiales de la « Current Bibliography for Aquatic Sciences and Fisheries », bibliografías especiales y trabajos relativos a los problemas de documentación.
MFS	Ediciones provisionales de los « Manuales FAO de la Ciencia de la Pesca ».

Algunos documentos tienen también otra identificación si, por ejemplo, son contribuciones a una reunión cuyos documentos han sido marcados con arreglo a otros sistemas.

FAO Fisheries Technical Paper No.26 Suppl. 1
(Distribution restricted)

FIb/T26 Suppl.1
MFS Sampling'

MANUAL OF SAMPLING AND STATISTICAL METHODS FOR FISHERIES BIOLOGY

Part II - STATISTICAL METHODS

Chapter No.5

COMPUTATIONS

by

K. P. Anderson
(Copenhagen, Denmark)

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 1965

24837/E

PREPARATION OF THIS PAPER

At the Symposium of ICES on Measurement of the abundance of fish populations, held at Madrid in 1963, the need was suggested for a series of "statistical leaflets" each of which would briefly describe a method useful in fisheries research, illustrated where possible by worked examples from fisheries biological studies. ICES subsequently requested FAO co-operation in this project, and a joint Editorial Group was established, composed of J.A. Pope (Aberdeen, Scotland), Convenor K.P. Anderson (Copenhagen, Denmark), J.A. Gulland (Lowestoft, England) and S.J. Holt (FAO). It was decided that a suitable publication vehicle for the "leaflets" would be the existing series FAO Manuals in Fisheries Science, and that each could conveniently be issued as a chapter in No. 3 of that series, the first four chapters of which, written by Gulland - a member of the Editorial Group - deal with sampling methods.

The Editorial Group subsequently drew up a plan for the first few statistical chapters and enlisted the help of other scientists to prepare them. The plan is outlined in a FAO internal report FB/R85 (October 1964) and in a report to the Comparative Fishing Committee of ICES made by Mr. Pope in Copenhagen, September 1964.

The Comparative Fishing Committee agreed to the procedure proposed by the Editorial Group for the revision of draft chapters. By this procedure the first drafts, after editing by the Editorial Group, are passed for review by members of the Comparative Fishing Committee and subsequently "adopted" by the Committee. After this they are issued as provisional versions in the original language in the series FAO Fisheries Technical Papers, and distributed more widely for comment, and trial use in Training Courses etc. Finally, after further revision they will be published in English, French and Spanish in the "Manuals" series.

PREPARATION DE CE DOCUMENT

Durant le Colloque de l'ICES sur l'Evaluation de l'abondance des populations de poissons, tenu à Madrid en 1963, il a été fait mention de la nécessité de créer une série de "Brochures Statistiques", dont chacune d'entre elles décrirait brièvement une méthode d'intérêt pour la recherche des pêches. Ces brochures seraient illustrées si possible par des exemples appliqués provenant d'études biologiques sur les pêches. L'ICES, en conséquence, a sollicité l'aide de la FAO pour ce projet et un Groupe Editorial mixte a été établi, comprenant J.A. Pope (Aberdeen, Ecosse), K.P. Anderson, Organisateur (Copenhague, Danemark), J.A. Gulland (Lowestoft, Angleterre) et S.J. Holt (FAO). Il a été décidé que le moyen le plus approprié à la parution de ces "brochures" était d'utiliser la série déjà existante de Manuel FAO de la Science des Pêches et que chaque brochure constituerait un chapitre dans le No. 3 de cette série. Mr. Gulland - membre du Groupe Editorial - serait chargé de la rédaction des quatre premiers chapitres qui traiteraient les méthodes d'échantillonnage.

Le Groupe Editorial a donc prévu un programme pour les premiers chapitres relatifs aux statistiques et a établi une liste des autres scientifiques qui aideront à leur préparation. Le programme est exposé dans le rapport interne de la FAO, FB/R85 (Octobre 1964) et dans un rapport pour le Comité comparatif des Pêches de l'ICES, présenté par Mr. Pope à Copenhague en Septembre 1964.

Le Comité comparatif des pêches a accepté la procédure proposée par le Groupe Editorial en ce qui concerne la révision du projet des chapitres. Selon cette procédure les premiers projets, après avoir été révisés par le Groupe Editorial, seront transmis aux membres du Comité comparatif des pêches afin d'être vérifiés et "adoptés" par ce dernier. Ensuite ils seront publiés provisoirement en langue originale dans les FAO Fisheries Technical Papers largement distribués pour commentaire et également utilisés, à titre d'essai, dans les cours d'entraînment etc. Après révision ultérieure, ils seront publiés en Anglais, Français et Espagnol dans les séries des "Manuels".

PREPARACION DE ESTE DOCUMENTO

En el Simposio del ICES sobre Medición de la abundancia de las poblaciones de peces, celebrado en Madrid en 1963, fué sugerida la necesidad de publicar una serie de "folletos estadísticos" cada uno de los cuales describiría brevemente un método válido para la investigación pesquera, ilustrado, cuando fuera posible, con ejemplos prácticos según la experiencia de organismos biológicos pesqueros. El ICES solicitó posteriormente la colaboración de la FAO en este proyecto y para ello se creó un Grupo Editorial conjunto, compuesto de J.A. Pope (Aberdeen, Escocia), convocador K.P. Anderson (Copenhage, Dinamarca), J.A. Gulland (Lowestoft, Inglaterra) y S.J. Holt (FAO). Se decidío que un medio conveniente para la publicación de estos "folletos" podría ser la ya existente serie de Manuales FAO Sobre Ciencias de la Pesca y que cada uno podría publicarse convenientemente como un capítulo en el No. 3 de dicha serie, estando los primeros cuatro capítulos, escritos por Gulland - miembro del Grupo Editorial - dedicados a los métodos de muestreo.

El Grupo Editorial preparó posteriormente un plan para los primeros capítulos estadísticos y buscó la ayuda de otros científicos para prepararlos. El plan se describe en un informe interno de la FAO, FB/R85 (Octubre 1964) y en el Informe al Comité de Pesca Comparada del ICES hecho por Mr. Pope en Copenhage, en Septiembre de 1964.

El Comité de Pesca Comparada aprobó el procedimiento propuesto por el Grupo Editorial para la revisión de los capítulos redactados. De este modo los primeros capítulos, después de corregidos por el Grupo Editorial se envian para su revisión a los miembros del Comité de Pesca Comparada y posteriormente son "adoptados" por el Comité. Después se publican como versiones provisionales en el idioma original de la serie de FAO Fisheries Technical Papers y se distribuyen ampliamente para obtener comentarios y para su empleo experimental en Centros de Capacitacíon etc. Por último, después de su revisión son publicados en inglés, francés y español en la serie de "Manuales".

Distribution

Author
FAO Fisheries Division
FAO Regional Fisheries Officers
FAO/EPTA Experts in Fisheries Biology
Biological Staff of UNSF Fisheries Projects
Selectors 2 and 5

"Current Bibliography" entry:

Anderson, K.P. (1965) 9-11020 MF
FAO Fish.tech.Pap., (26) Suppl.1:25 p.
Manual of sampling and statistical methods for fisheries biology. Part 4 Statistical methods. Chapter 5 Computations.

Gives methods for handling inaccurate figures, calculations with inaccurate data, computations and reliability, effect of grouping, functions, errors in solution of simultaneous equations, and the statement of error.
Co 62-04022

C O N T E N T S

Chapter	5	COMPUTATIONS	
	5.1	Rounding off and inexactness	5:1
	5.2	Calculations with inaccurate data	5:3
	5.3	Computations and reliability	5:6
	5.4	The effect of grouping	5:13
	5.5	More complicated functions	5:18
	5.6	Errors in the solution of simultaneous equations	5:21
	5.7	The statement of error	5:23

5. COMPUTATIONS by K. P. Anderson

This chapter deals with the handling of inaccurate figures in computations and some related problems. The underlying mathematical theory is dealt with rather loosely but all results are illustrated by numerical examples. In most cases the examples use real data from Danish investigations and I am very grateful to Mag. K. Popp Madsen for putting these at my disposal.

5.1. Rounding off and inexactness

We start by defining inexactness.

Definition: If a figure b is given in the form

$$b = a \times 10^n$$

where $1 \leq a < 10$

and if a has N figures after the decimal point then b is stated with an inexactness of

$$5 \times 10^{-(N+1)} \times 10^n = 5 \times 10^{n-N-1}$$

Example 5.1.1.

If the total length l of a fish is given as

$$l = 2.43 \times 10 \text{ cm}$$

then l is stated with an inexactness of

$$5 \times 10^{1-2-1} \text{cm} = 5 \times 10^{-2} \text{cm}$$

Example 5.1.2.

The total catch of a ship is given as 5.1 x 10 tons.

The total catch is stated with an inexactness of

$$5 \times 10^{1-1-1} = 5 \times 10^{-1} \text{ tons}$$

The definition implies a distinction between, for example, the two figures 3.1 and 3.10. The first figure is stated with an inexactness of $5 \times 10^{0-1-1} = 5 \times 10^{-2}$ whereas the second one is stated with an inexactness of $5 \times 10^{0-2-1} = 5 \times 10^{-3}$.

If figures are written in ways other than that indicated in the definition it is first necessary to convert to the form given there and then apply the rule.

Example 5.1.3.

The area of a lake is 243.1 km^2. The area can be written as $2.431 \times 10^2 km^2$ and the area is thus stated with an inexactness of $5 \times 10^{2-3-1}\ km^2 = 5 \times 10^{-2}\ km^2$.

Example 5.1.4.

The length of a diatom is 0.017 mm.

The length can be written as 1.7×10^{-2} mm and is stated with an inexactness of $5 \times 10^{-2-1-1} = 5 \times 10^{-4}$ mm.

Difficulties arise with figures such as 21500. Should this be written as 2.15×10^4 or 2.150×10^4 or perhaps as 2.1500×10^4? The three possibilities are stated with different inexactness and this sort of ambiguity should always be avoided. The figure should be written as 2.150×10^4 or 21.50×10^3 or 215.0×10^2 if it is intended to imply that the inexactness is $5 \times 10^0 = 5$.

The definition is derived from the process of rounding off. Consider, for example, the following number

$$\log_{10} 2 = 0.3010299956....$$

Rounding off to 8 decimals, 5 decimals and 3 decimals give, respectively,

$$\log_{10} 2 = 0.30103000$$

$$\log_{10} 2 = 0.30103$$

$$\log_{10} 2 = 0.301$$

With our definition we see that these figures are stated with an inexactness of 5×10^{-9}, 5×10^{-6} and 5×10^{-4} respectively. The point is that <u>the differences between the rounded-off values and the true value are never greater than the inexactness</u> if the rounding off is done properly.

Rounding off to <u>m</u> decimals should be done as follows.

If the difference between the true value and the figure with the same first <u>m</u> decimals and zeroes in the last (infinitely many) decimals is $< 5 \times 10^{-(m+1)}$ the m'th(last) decimal in the rounded off figure equals the m'th decimal in the true value. If the difference is $> 5 \times 10^{-(m+1)}$ the m'th (last) decimal in the rounded off figure is the decimal of the true value raised by 1.

The rule states nothing about a number such as 1.2500000.... If we want to round off to one decimal place both 1.2 and 1.3 differ from the true value by the same amount, viz. 5×10^{-2}. To preserve the randomness of the rounding-off error this limiting case is handled by the following rule.

If the difference between the true value and the rounded off value equals $5 \times 10^{-(m+1)}$ the m'th decimal is unchanged if even and raised by 1 if uneven.

The figure 1.25000... is then 1.2 when rounded off to 1 decimal, but 1.35000..... becomes 1.4.

If one really knew figures with an infinite number of decimals, the rules stated above would always give differences between true values and rounded off values that were no greater than $5 \times 10^{-(m+1)}$ where m is the number of decimals in the rounded off figure. The real situation is, however, that in most cases only a very few decimals are known beyond the rounded off one. In many cases only one further decimal is known. The rules are applied in this situation too, but this means that the difference between true value and rounded off value can be as great as $5.5 \times 10^{-(m+1)}$.

5.2. Calculations with inaccurate data

In this section we show how to evaluate the errors in expressions from a knowledge of the errors in the constituents.

We require to introduce the concept of the modulus of a number. If $\underline{a}$ is any number, the modulus of $\underline{a}$ (written $|\underline{a}|$ and read as "mod $\underline{a}$") is the value of the number ignoring sign. That is, if $\underline{a}$ is positive $|\underline{a}| = \underline{a}$ while if $\underline{a}$ is negative $|\underline{a}| = -\underline{a}$. For example, $|6| = 6$, and $|-6| = 6$.

We will also require the following results which we state without proof.

(a) $|a + b| \leq |a| + |b|$

(b) $|a + b + c + \ldots| \leq |a| + |b| + |c| + \ldots$

(c) $|a - b| \leq |a| + |b|$

(d) $|a \pm b \pm c \pm \ldots| \leq |a| + |b| + |c| + \ldots$

(e) $|a \times b \times c \times \ldots| = |a| \cdot |b| \cdot |c| \ldots$

We next require a definition

Definition: If α is the exact value of a quantity and $\underline{a}$ an approximate value, so that

$$\underline{a} = \alpha + e$$

(approximate value) = (exact value) plus (error)

the absolute* error of $\underline{a}$, the relative error of $\underline{a}$ and the correction ($\underline{c}$) are defined as

$$\begin{aligned} \text{absolute error} &= a - \alpha = \Delta a \\ \text{relative error} &= \Delta a/\alpha \\ \text{correction} &= c = \alpha - a \end{aligned}$$

We shall work with absolute and relative errors only. Nothing definite will be supposed about the origin of the errors, but one common source is of course rounding off. If a figure is rounded off the absolute error is smaller than or equal to the inexactness the figure is stated with. It is clear that a rather good measure of the absolute error of rounded off figures is the number of correct figures after the decimal point. A comparable measure of the relative error of a rounded off figure is the number of significant figures, which is the number of decimals when the figure is written as

$$a \times 10^n \quad \text{with } 0.1 \leq a < 1$$

Now, let x and y be two inaccurate figures and let

$$\begin{aligned} x &= \xi + \Delta x \\ y &= \eta + \Delta y \end{aligned}$$

where ξ and η are the true values of x and y respectively.

The sum of these inaccurate figures is

$$x + y = (\xi + \Delta x) + (\eta + \Delta y) = (\xi + \eta) + (\Delta x + \Delta y)$$

As $\xi + \eta$ is the true value of the sum we see that

$$\Delta(x + y) = \Delta x + \Delta y$$

or, in words, the absolute error of a sum is the sum of the absolute errors of the addends.

For the product we find

$$\begin{aligned} xy = (\xi + \Delta x)(\eta + \Delta y) &= \xi\eta + \eta\Delta x + \xi\Delta y + \Delta x \cdot \Delta y \\ &\doteq \xi\eta + \eta\Delta x + \xi\Delta y \end{aligned}$$

if Δx and Δy are small compared with ξ and η (i.e. we drop the term $\Delta x \cdot \Delta y$). As the true value of the product is $\xi\eta$ this gives:

$$\Delta(xy) \sim \eta\Delta x + \xi\Delta y.$$

---

*Many text books use the word absolute synonymously with modulus as defined here. We will, however, strictly adhere to the definitions we have given.

The relative error of the product is

$$\frac{\Delta(xy)}{\xi\eta} \sim \frac{\Delta x}{\xi} + \frac{\Delta y}{\eta}$$

or, in words, the relative error of a product is (nearly) the sum of the relative errors of the factors.

The difference has the following absolute error

$$\Delta(x - y) = \Delta x - \Delta y$$

whereas the quotient gives

$$\Delta\left(\frac{x}{y}\right) = \left\{\frac{x}{y} - \frac{\xi}{\eta}\right\} = \frac{\xi + \Delta x}{\eta + \Delta y} - \frac{\xi}{\eta}$$
$$= \frac{\xi\eta + \eta\Delta x - \xi\eta - \xi\Delta y}{\eta(\eta + \Delta y)} \sim \frac{\Delta x}{\eta} - \frac{\xi\Delta y}{\eta^2}$$

if Δy is small compared with η.

In terms of relative errors this is

$$\frac{\Delta\left(\frac{x}{y}\right)}{\xi/\eta} \sim \frac{\Delta x}{\xi} - \frac{\Delta y}{\eta}$$

One use of these simple formulae is the determination of the maximum error of a calculated quantity if the maximum absolute errors of the components are known.

Example 5.2.1

The catch of a cutter is specified as:

Weight of small plaice	s = 11 t,	$\lvert\Delta s\rvert \leq 1$ ton
- - medium -	m = 19 t,	$\lvert\Delta m\rvert \leq 2$ tons
- - large -	l = 5 t,	$\lvert\Delta l\rvert \leq 0.5$ ton
- - by catch	b = 7 t,	$\lvert\Delta b\rvert \leq 1$ ton

What is the maximum absolute and relative error of the weight of the total catch?

The total catch is

$$W = s + m + l + b = 42 \text{ tons}$$

For the absolute error we find

$$\Delta W = \Delta(s + m + l + b) = \Delta s + \Delta m + \Delta l + \Delta b \leq 4.5 \text{ tons}$$

Now we wish to find the maximum value that ΔW can take, knowing that the error in s does not exceed 1 ton, etc. Normally the signs of the errors are not known so when estimating the maximum absolute error we combine the maximum errors of the constituents in the most unfavourable way. The modulus of the absolute error of W is

$$|\Delta W| = |\Delta s + \Delta m + \Delta l + \Delta b|$$
$$\leq |\Delta s| + |\Delta m| + |\Delta l| + |\Delta b|$$

from inequality (b) above.

Inserting the values of $|\Delta s|$ etc. we find

$$|\Delta W| \leq (1 + 2 + 0.5 + 1) = 4.5 \text{ tons.}$$

The relative error is $\Delta W/T$ where T is the unknown true total weight. Since W is approximately equal to T we may write $\Delta W/T \sim \Delta W/W$ and the maximum modulus value of the relative error is

$$\left|\frac{\Delta W}{W}\right| \leq \frac{|\Delta s| + |\Delta m| + |\Delta l| + |\Delta b|}{|W|} = \frac{4.5}{42} = 0.11.$$

That is, the relative error in W is not greater than 11%.

5.3. Computations and reliability

So far we have written about unavoidable errors in calculations caused by the necessity of working with inaccurate data. The next problem to study is that of the calculation itself. It will be supposed that the reader is accustomed to the use of slide-rules and electrical calculating machines as well as tables.

The first step is to determine the number of figures after the decimal point which are necessary to give the required accuracy in the calculated result. This is normally done with reference to maximum errors as dealt with in the previous sections. By doing this at the beginning of a calculation one avoids using either too many or too few decimals, the former leading to unnecessary work and the latter to useless results.

When the number of figures after the decimal point has been determined, the type of calculating aid is chosen. If an accuracy of 1% is needed a slide-rule is suitable for most purposes, but if higher accuracy is needed electrical calculators with registers of suitable capacity have to be used.

The real trouble from now on in the calculations arises from mistakes. All calculations should be arranged in such a way that mistakes are avoided as far as possible. Mistakes cannot, of course, be completely avoided, and it is therefore necessary to have checks that can be applied

step by step in such a way that mistakes are discovered immediately and not carried forward to the final answer.

Some common causes of mistakes are

a) Misreading of tables
b) Mistakes in signs
c) Misplacing decimal points
d) Interchanges of digits

The adoption of an orderly scheme of computing, embodying checks, will greatly minimize the risk of computational errors.

When checking identities rather than repetitions should be employed as it is a common experience that the same errors can be committed in repetitions.

Some procedures will now be demonstrated by means of examples.

Example 5.3.1

Eleven herring have the following weights in grams

77.8 36.9 12.7 39.1 60.2 46.5 75.9 41.2 41.4 88.0 82.7

Find the arithmetic mean and the variance.

The mean and variance are defined as

$$\text{Mean} = \bar{w} = \frac{\sum w_i}{n} \quad (n = \text{number of fish})$$

$$\text{Variance} = s^2 = \frac{\sum (w_i - \bar{w})^2}{n-1} \equiv \frac{\sum w_i^2 - (\sum w_i)^2/n}{n-1}$$

As regards the variance the expression on the extreme right is normally used as many desk calculators give $\sum w_i^2$ and $\sum w_i$ in one operation. Calculations give

$$\sum w_i = 602.4$$
$$\sum w_i^2 = 38646.34$$

As a check we use the identities

$$\sum_{i=1}^{n} (x_i + u) = \sum x_i + n\,u$$

$$\sum_{i=1}^{n} (x_i + u)^2 = \sum x_i^2 + 2\,u \sum x_i + n\,u^2$$

If we put $u = 0.3$, for example, we get the new "weights" w'

78.1 37.2 13.0 39.4 60.5 46.8 76.2 41.5 41.7 88.3 83.0

and

$$\sum w' = 605.7$$
$$\sum w'^2 = 39008.77$$

The checking is now done by calculating

$$\sum w_i + (11 \times 0.3) = 602.4 + 3.3 = 605.7$$
$$\sum w_i^2 + (2 \times 0.3 \sum w_i) + (11 \times 0.3^2) = 38646.34 + 361.44 + 0.99 = 39008.77$$

The check is not completely safe but it is much better than a repetition.

Let us estimate the maximum error of $\sum w$ and $\sum w^2 - (\sum w)^2/n$. We have

$$|\Delta(\sum w)| \leq \sum |\Delta w_i| = n\,|\Delta w|$$
$$|\Delta(\sum w^2 - (\sum w)^2/n)| \leq \sum \left\{ \left| 2w - \frac{2\sum w}{n} \right| \cdot |\Delta w| \right\}^*$$

If we suppose $|\Delta w| \leq 0.05$ g for all fish we get

$$|\Delta(\sum w)| \leq 11 \times 0.05 = 0.55$$

and

$$|\Delta(\sum w^2 - (\sum w)^2/n)| \leq 2 \times 221.56 \times 0.05 = 22$$

where each final figure is given with 2 significant digits.

The mean and variance should thus be given as

$$\bar{w} = \frac{602.4}{11} = 54.764 \pm 0.050$$
$$s^2 = \frac{38646.34 - \frac{(602.4)^2}{11}}{10} = 565.7 \pm 2.2$$

If in the calculations the figures had first been rounded off to the nearest gram we would have

$$|\Delta w| \leq 0.5$$

and

$$|\Delta(\sum w)| \leq 5.5$$
$$|\Delta(\sum w^2 - (\sum w)^2/n)| \leq 222$$

*

See Section 5.5 for proof of these results.

and the mean and variance calculated from the rounded off figures should have been written

$$\bar{w} = \frac{602}{11} = 54.73 \pm 0.50$$

$$s^2 = \frac{38630 - \frac{602^2}{11}}{10} = 568 \pm 22$$

Which of the two accuracies should be preferred cannot be decided on the basis of the theory outlined here.

The inaccuracies found are actually calculating accuracies originating from rounding off, and we see that if we want

$$|\Delta \bar{w}| \leq 0.050$$
$$|\Delta s^2| \leq 5.0$$

we should use all given decimals whereas for

$$|\Delta \bar{w}| \leq 0.50$$
$$|\Delta s^2| \leq 50$$

whole gram figures would do. To follow the scheme set up at the beginning of this paragraph we should have started by finding the number of decimals sufficient for the desired accuracy in the calculated results. In this case (as in many others) this is rather impractical. The amount of calculation is about the same whether the given figures or figures rounded off to the nearest gram are used. In many cases it is therefore sound practice to use the original data and then do the rounding off in the results. It is always possible to round off if too many figures have been carried, whereas it will be impossible to achieve the desired accuracy if too few figures have been carried.

For grouped data another check, namely the use of factorial moments, can be used with advantage. The definition and theory of factorial moments will not be dealt with here, but the method will be illustrated by an example.

Example 5.3.2

The vertebral counts of a herring sample are given in the first two columns of the following table.

V.S.	frequency	I	II	III
59	2	2	2	2
58	34	36	38	40
57	340	376 = a	414 = d	454 = f
56	284 = b			
55	19	22 = c	26 = e	31 = g
54	2	3	4	5
53	1	1	1	1

Find the mean and variance.

The usual procedure gives

$$\sum(\text{V.S.})^2 = 2182738 \qquad \sum \text{V.S.} = 38580 \qquad n = 682$$

Checking gives

$$\sum(\text{V.S.} + 1) = 39262 = 38580 + 682$$
$$\sum(\text{V.S.} + 1)^2 = 2260580 = 2182738 + (2 \times 38580) + 682$$

To check using factorial moments the procedure is as follows. First a "working mean" is chosen. In the example the working mean is chosen as 56. Then the frequencies above 56 are accumulated upwards:

$$2,\ 2 + 34 = 36,\ 2 + 34 + 340 = 376$$
$$1,\ 1 + 3 = 4,\ 1 + 3 + 22 = 26$$

This gives us column I.

Column I is then accumulated in the same way

$$2,\ 2 + 36 = 38,\ \ 2 + 36 + 376 = 414$$
$$1,\ 1 + 3 = 4,\ \ 1 + 3 + 22 = 26$$

This gives column II, which is also accumulated:

$$2,\ 2 + 38 = 40,\ \ 2 + 38 + 414 = 454$$
$$1,\ 1 + 4 = 5,\ \ 1 + 4 + 26 = 31$$

these figures being shown in column III.

It can now be shown that

$$\sum \text{V.S.} = n\ (\text{working mean}) + (d - e)$$
$$\sum(\text{V.S.})^2 = 2\ (f + g) - (d + e) + 2\ (\textstyle\sum \text{V.S.})\ (\text{working mean}) - n(\text{working mean})^2.$$

For the example the formulae give

$$682 \times 56 + (414 - 26) = 38580$$
$$2(454 + 31) - (414 + 26) + 2 \times 38580 \times 56 - 682 \times 56^2 = 2182738$$

If the grouping interval has a length h the formulae become

$$\sum x = n(\text{working mean}) + h(d - e)$$
$$\sum x^2 = (2\ (f + g\ ) - (d + e)\ )\ h^2 + 2(\text{working mean}) \sum x - n(\text{working mean})^2$$

The mean and variance can then be determined from the formulae

$$\bar{x} = (\text{working mean}) + \frac{h(d - e)}{n}$$

$$s^2 = \frac{2(f + g) - (d + e) - (d - e)^2/n}{n - 1} \cdot h^2$$

These formulae are in many cases practical to use and a check can be obtained by using different working means.

All methods give for this example

$$\overline{V.S.} = 56.569$$

$$s^2 = 0.4541$$

and in this case only the final rounding off gives any inaccuracy. As in the former example the question of the correct number of digits is not solved by this but has to be decided on the basis of statistical theorems.

In addition to sums and sums of squares, product sums are also very important in statistical calculations. A check for the latter is provided by the identity

$$\sum(x + a)(y + b) = \sum x.y + b\sum x + a\sum y + nab$$

We finish this section with some words about square roots, which also play an important role in statistics. When calculating square roots on desk computers iterative methods are convenient. The following method gives accurate roots with great speed.

Let N be the number whose square root is desired and let A_1 be an approximation to $\sqrt{N}$.

Calculate $N/A_1 = A_1'$ (say). A second approximation is given by

$$A_2 = \tfrac{1}{2}(A_1 + A_1')$$

Next calculate $N/A_2 = A_2'$ (say). A third approximation is given by

$$A_3 = \tfrac{1}{2}(A_2 + A_2')$$

Continue this process until the desired accuracy is obtained.

Example 5.3.3

Find the square root of 2 with 6 decimals.

As first approximation we take

$$A_1 = 1.4000000 \quad \text{(seven decimals)}$$

which means that

$$A_1' = 2/1.4000000 = 1.4285714$$

and a second approximation is

$$A_2 = 1/2(A_1 + A_1') = 1.4142857$$

The third approximation is found from

$$A_2' = 2/A_2 = 1.4141414$$

and is $A_3 = 1/2(A_2 + A_2') = 1.4142135$

For the fourth approximation we get

$$A_3' = 2/A_2 = 1.4142136$$

$$A_4 = 1/2(A_3 + A_3') = 1.4142136$$

The difference between A_3 and A_4 is negligible so that we may terminate the process with the result that

$$\sqrt{2} = 1.414214$$

It is not very important that A_1 be close to $\sqrt{N}$. The iteration will lead to $\sqrt{N}$ rapidly anyway.

Example 5.3.4

Find $\sqrt{7.113214}$ with 6 decimals.

$A_1 = 1.0000000$ $A_1' = 7.1132140$

$A_2 = 4.0566070$ $A_2' = 1.7534776$

$A_3 = 2.9050478$ $A_3' = 2.4485704$

$A_4 = 2.6768091$ $A_4' = 2.6573483$

$A_5 = 2.6670787$ $A_5' = 2.6670432$

$A_6 = 2.6670609$ $A_6' = 2.6670610$

which gives

$$\sqrt{7.113214} = 2.667061$$

The maximum error of a square root estimated from the maximum error of the radical is

$$|\Delta x^{\frac{1}{2}}| \leq \frac{1}{2x^{\frac{1}{2}}} |\Delta x|$$

and the relative error is

$$\left|\frac{\Delta x^{\frac{1}{2}}}{x^{\frac{1}{2}}}\right| \leq \frac{1}{2}\left|\frac{\Delta x}{x}\right|$$

That is:

The relative error of a square root is half the relative error of the radical.

5.4. The effect of grouping

It will often be necessary or convenient to work with grouped data. Information is clearly lost when data are grouped and in this section the effect of grouping on the mean and variance is studied. The statistical theory of grouping is too complicated to be dealt with here, and we shall merely illustrate by means of an example.

Example 5.4.1

Length measurements to the millimetre below of 150 herrings gave the results given in table 5.4.1

Table 5.4.1 (Length in cm)

17.9	16.1	13.2	15.3	16.9	15.9	15.9	16.4	16.7	15.9
15.5	16.8	14.8	15.7	15.1	13.3	15.6	15.3	14.2	13.5
17.2	15.1	16.0	16.3	12.0	15.7	15.6	16.3	14.8	15.8
14.6	13.9	15.9	16.5	16.1	17.2	11.7	15.3	16.3	12.3
15.5	15.2	16.2	12.8	14.5	16.1	16.4	12.5	14.7	15.7
16.4	15.8	16.0	16.1	16.4	14.1	16.2	16.7	15.1	15.0
15.7	14.2	13.2	13.9	16.1	15.1	17.1	17.1	14.2	14.3
16.3	15.0	14.2	14.8	17.1	15.0	15.1	16.3	15.4	14.8
15.5	16.3	16.2	15.7	15.7	14.8	12.9	14.0	15.7	13.6
15.8	16.2	15.6	16.6	11.9	13.7	16.9	15.6	16.5	17.8
16.3	15.8	15.0	15.3	15.4	18.1	17.9	13.7	15.4	14.2
15.7	14.5	16.3	16.4	13.3	16.0	12.9	15.2	14.9	15.3
16.9	14.1	16.0	16.0	16.0	16.8	15.7	12.8	14.6	17.4
14.6	12.7	16.8	15.9	14.7	14.8	12.9	14.9	14.0	14.7
16.9	14.9	16.8	15.2	14.6	12.3	16.1	15.8	18.3	14.3

The data in table 5.4.1 are also grouped, the grouping interval being 1 mm, but as this is a small interval we will refer to them as the ungrouped data.

For the ungrouped data we find

$$\sum l = 2299.0 \qquad \sum l^2 = 35501.64 \qquad n = 150$$

$$\bar{l} = \frac{2299.0}{150} + 0.050 = 15.38$$

$$s^2 = 1.783$$

When the data are grouped into larger intervals this process is only unique if the centres of the series of grouping intervals are given. There are for the given data five different semi-centimetre groupings, viz.

1) (···, 11.75, 12.25, 12.75, 13.25, ···)
2) (···, 11.65, 12.15, 12.65, 13.15, ···)
3) (···, 11.55, 12.05, 12.55, 13.05, ···)
4) (···, 11.45, 11.95, 12.45, 12.95, ···)
5) (···, 11.35, 11.85, 12.35, 12.85 ···)

where the figures indicate the centres of the grouping intervals.

The five groupings transform table 5.4.1 into table 5.4.2.

Table 5.4.2

length	no.	length	no.	length	no.	length	no.	length	no.
18.75		18.65		18.55	1	18.45	1	18.35	2
18.25	2	18.15	4	18.05	4	17.95	4	17.85	3
17.75	3	17.65	2	17.55	1	17.45	3	17.35	6
17.25	6	17.15	9	17.05	13	16.95	13	16.85	11
16.75	13	16.65	14	16.55	18	16.45	30	16.35	25
16.25	29	16.15	29	16.05	26	15.95	31	15.85	29
15.75	26	15.65	24	15.55	24	15.45	18	15.35	19
15.25	20	15.15	20	15.05	21	14.95	21	14.85	20
14.75	18	14.65	15	14.55	11	14.45	13	14.35	11
14.25	11	14.15	13	14.05	11	13.95	8	13.85	7
13.75	6	13.65	4	13.55	6	13.45	6	13.35	5
13.25	4	13.15	7	13.05	7	12.95	6	12.85	6
12.75	7	12.65	4	12.55	4	12.45	3	12.35	3
12.25	3	12.25	4	12.05	2	11.95	3	11.85	3
11.75	2	11.65	1	11.55	1	11.45		11.35	
(1)		(2)		(3)		(4)		(5)	

Calculations utilizing factorial moments give

$$\bar{I} = 15.25 + \frac{174 - 141}{2 \times 150} = 15.36$$

$$s^2 = \frac{2(338 + 351) - (174 + 141) - (174 - 141)^2/150}{4 \times 149} = 1.771 \qquad (1)$$

$$\bar{I} = 15.15 + \frac{194 - 132}{2 \times 150} = 15.36$$

$$s^2 = \frac{2(399 + 320) - (194 + 132) - (194 - 132)^2/150}{4 \times 149} = 1.823 \qquad (2)$$

$$\bar{I} = 15.55 + \frac{131 - 181}{2 \times 150} = 15.38$$

$$s^2 = \frac{2(249 + 461) - (131 + 181) - (131 - 181)^2/150}{4 \times 149} = 1.831 \qquad (3)$$

$$\bar{l} = 15.45 + \frac{148 - 164}{2 \times 150} = 15.40 \qquad (4)$$

$$s^2 = \frac{2(280 + 405) - (148 + 164) - (148 - 164)^2/150}{4 \times 149} = 1.772$$

$$\bar{l} = 15.35 + \frac{163 - 152}{2 \times 150} = 15.39 \qquad (5)$$

$$s^2 = \frac{2(317 + 382) - (163 + 152) - (163 - 152)^2/150}{4 \times 149} = 1.816$$

If we take the mean of these different means and variances we get

$$\hat{l}_{\frac{1}{2}} = \text{Mean } \bar{l} = 15.38$$

$$\hat{s}^2_{\frac{1}{2}} = \text{Mean } s^2 = 1.803$$

The result is that

$$\hat{l}_{\frac{1}{2}} = \bar{l} \text{ for the ungrouped data}$$

$$\hat{s}^2_{\frac{1}{2}} > s^2 \text{ for the ungrouped data}$$

Grouping the data into 1 cm intervals gives table 5.4.3

Table 5.4.3

length	no.	length	no.	length	no.	length	no.	length	no.
19.5		19.4		19.3		19.2		19.1	
18.5	2	18.4	4	18.3	5	18.2	5	18.1	5
17.5	9	17.4	11	17.3	14	17.2	16	17.1	17
16.5	42	16.4	43	16.3	44	16.2	51	16.1	54
15.5	46	15.4	44	15.3	45	15.2	39	15.1	39
14.5	29	14.4	28	14.3	22	14.2	21	14.1	18
13.5	10	13.4	11	13.3	13	13.2	12	13.1	11
12.5	10	12.4	8	12.3	6	12.2	6	12.1	6
11.5	2	11.4	1	11.3	1	11.2		11.1	
10.5		10.4		10.3		10.2		10.1	
(1)		(2)		(3)		(4)		(5)	
19.0		18.9		18.8	1	18.7	1	18.6	2
18.0	5	17.9	6	17.8	5	17.7	7	17.6	9
17.0	19	16.9	23	16.8	31	16.7	33	16.6	36
16.0	55	15.9	53	15.8	50	15.7	49	15.6	48
15.0	38	14.9	35	14.8	32	14.7	34	14.6	31
14.0	17	13.9	17	13.8	17	13.7	14	13.6	12
13.0	11	12.9	11	12.8	11	12.7	9	12.6	9
12.0	5	11.9	5	11.8	3	11.7	3	11.6	3
11.0		10.9		10.8		10.7		10.6	
10.0		9.9		9.8		9.7		9.6	
(6)		(7)		(8)		(9)		(10)	

Table 5.4.3 gives the following means and variances

	$\bar{l}$	s^2		$\bar{l}$	s^2
(1)	15.36	1.906	(6)	15.36	1.762
(2)	15.39	1.899	(7)	15.32	1.869
(3)	15.41	1.927	(8)	15.35	1.860
(4)	15.43	1.844	(9)	15.36	1.823
(5)	15.40	1.809	(10)	15.37	1.948

The mean mean and the mean variance are

$$\text{mean } \bar{l} = \hat{l}_1 = 15.38$$

$$\text{mean } s^2 = \hat{s}^2_1 = 1.865$$

and we find

$$\hat{l}_1 = \bar{l} \text{ (ungrouped data)}$$

$$\hat{s}^2_1 > \hat{s}^2_{\frac{1}{2}} > s^2 \text{ (ungrouped data)}$$

Finally grouping into intervals of 2 cm gives table 5.4.4

Table 5.4.4

length	no.	length	no.	length	no.	length	no.	length	no.
21.0		20.9		20.8		20.7		20.6	
19.0	2	18.9	4	18.8	5	18.7	5	18.6	5
17.0	51	16.9	54	16.8	58	16.7	67	16.6	71
15.0	75	14.9	72	14.8	67	14.7	60	14.6	57
13.0	20	12.9	19	12.8	19	12.7	18	12.6	17
11.0	2	10.9	1	10.8	1	10.7		10.6	
9.0		8.9		8.8		8.7		8.6	
	(1)		(2)		(3)		(4)		(5)
20.5		20.4		20.3		20.2		20.1	
18.5	5	18.4	6	18.3	6	18.2	8	18.1	11
16.5	74	16.4	76	16.3	81	16.2	82	16.1	84
14.5	55	14.4	52	14.3	49	14.2	48	14.1	43
12.5	16	12.4	16	12.3	14	12.2	12	12.1	12
10.5		10.4		10.3		10.2		10.1	
8.5		8.4		8.3		8.2		8.1	
	(6)		(7)		(8)		(9)		(10)
20.0		19.9		19.8		19.7		19.6	
18.0	11	17.9	15	17.8	19	17.7	21	17.6	22
16.0	88	15.9	87	15.8	89	15.7	90	15.6	93
14.0	39	13.9	39	13.8	35	13.7	33	13.6	29
12.0	12	11.9	9	11.8	7	11.7	6	11.6	6
10.0		9.9		9.8		9.7		9.6	
8.0		7.9		7.8		7.7		7.6	
	(11)		(12)		(13)		(14)		(15)
19.5		19.4		19.3	1	19.2	1	19.1	2
17.5	24	17.4	29	17.3	36	17.2	40	17.1	45
15.5	93	15.4	88	15.3	82	15.2	83	15.1	79
13.5	28	13.4	28	13.3	28	13.2	23	13.1	21
11.5	5	11.4	5	11.3	3	11.2	3	11.1	3
9.5		9.4		9.3		9.2		9.1	
7.5		7.4		7.3		7.2		7.1	
	(16)		(17)		(18)		(19)		(20)

Calculating variances and means as before gives:

	$\bar{l}$	s^2		$\bar{l}$	s^2
(1)	15.41	2.164	(11)	15.31	2.147
(2)	15.45	2.196	(12)	15.34	2.100
(3)	15.43	2.316	(13)	15.40	2.040
(4)	15.49	2.196	(14)	15.38	1.991
(5)	15.45	2.166	(15)	15.35	1.949
(6)	15.41	2.125	(16)	15.31	1.898
(7)	15.36	2.186	(17)	15.28	2.053
(8)	15.35	2.078	(18)	15.35	2.145
(9)	15.35	2.059	(19)	15.37	2.091
(10)	15.35	2.177	(20)	15.39	2.222

mean $\bar{l} = \hat{l}_2 = 15.38 = \bar{l}$ (ungrouped data)

mean $s^2 = \hat{s}_2^2 = 2.115 > \hat{s}_1^2 > \hat{s}_{\frac{1}{2}}^2 > s^2$ (ungrouped data)

When ungrouped data are grouped one can on the basis of theoretical statistics prove that the grouped mean is equal to the ungrouped mean on the average and that the grouped variance is, on the average, equal to the ungrouped variance plus $h^2/12$ where h is the length of the grouping interval (Sheppard's corrections). The above calculation is a demonstration of this theorem and the following table shows that millimetre groups in this example can be considered as ungrouped.

h	$h^2/12$	$s^2 + h^2/12$	s_2^2
$\frac{1}{2}$	0.021	1.804	1.803
1	0.083	1.866	1.865
2	0.333	2.116	2.115

The process of grouping can be looked upon as a process of making the data more inaccurate, a phenomenon that shows up as an increase in the calculated variance. This can be taken as a loss of information, and it is clear that the smaller the grouping interval the smaller the loss of information. On the other hand the larger the grouping interval the larger the saving in labour and time. The chosen grouping interval has in this way to be a compromise between the two points of view. But there is another aspect of importance and this is clearly shown in the example. The variation in the means and variances for the same data and grouping interval but for different interval centres increases rapidly with the size of the grouping interval. This has a direct bearing on the reliability of any statistical tests carried out with grouped data. We will not go further into this problem, but only state that it is the size of the variance that determines the safe upper limit of the grouping interval as regards statistical tests. A useful rule for determining the safe length of grouping intervals is

The square of the length of the grouping interval should be smaller than one quarter of the variance, or in symbols

$$h^2 < s^2/4$$

Example 5.4.2

For Example 5.4.1 $s^2 = 1.8$ and the rule states that h^2 should be smaller than 0.45. The maximum value of h is in this way about 0.7 and this means that the grouping interval used in practice should be 0.5 cm.

5.5 More complicated functions

In this section we extend the results given in Section 5.2 to more complicated expressions. The ability to differentiate simple functions is necessary for a proper understanding of what follows.

Let u be a function of quantities $x, y, z, \ldots$. For example, if u is the condition factor of a fish, then u is a function of the weight x and the length of y, the form of the function being $u = x/y^3$.

If $X, Y, Z, \ldots$ are the true values of $x, y, z, \ldots$, that is, if $x = X + \Delta x$, $y = Y + \Delta y$, $z = Z + \Delta z, \ldots$ we may write

$$u = f(X + \Delta x, Y + \Delta y, Z + \Delta z, \ldots)$$

$$= f(X, Y, Z, \ldots) + \left(\frac{\partial f}{\partial x}\right)_{x_1 y_1 z_1 \ldots} \Delta x + \left(\frac{\partial f}{\partial y}\right)_{x_1 y_1 z_1 \ldots} \Delta y + \left(\frac{\partial f}{\partial z}\right)_{x_1 y_1 z_1 \ldots} \Delta z + \ldots$$

where x_1, y_1, z_1 are values in the intervals (X,x), (Y,y), (Z,z), etc. Then

$$u - f(X, Y, Z) = \Delta u$$

$$= \left(\frac{\partial f}{\partial x}\right) \Delta x + \left(\frac{\partial f}{\partial y}\right) \Delta y + \left(\frac{\partial f}{\partial z}\right) \Delta z + \ldots \qquad \ldots. (5.5.1)$$

where the various first-order partial derivatives are understood to be evaluated at $x = x_1, y = y_1, z = z_1 \ldots$.

Hence

$$|\Delta u| = \left| \left(\frac{\partial f}{\partial x}\right) \Delta x + \left(\frac{\partial f}{\partial y}\right) \Delta y + \left(\frac{\partial f}{\partial z}\right) \Delta z + \ldots \right|$$

$$\leq \left|\frac{\partial f}{\partial x}\right| |\Delta x| + \left|\frac{\partial f}{\partial y}\right| |\Delta y| + \left|\frac{\partial f}{\partial z}\right| |\Delta z| + \ldots$$

This last expression gives the maximum value of the absolute error in the calculated quantity. When estimating maximum errors there is no need for great accuracy; two significant digits are sufficient.

Example 5.5.1

The total length of a herring is given as $l = 32$ cm, $|\Delta l| \leq 1$ cm and the total weight as $w = 300$ gm, $|\Delta w| \leq 5$ gm. Find the condition factor q and estimate its maximum error.

Since $q = w/l^3$ we have

$$\frac{\partial f}{\partial w} = \frac{1}{l^3} \qquad \text{and} \qquad \frac{\partial f}{\partial l} = -\frac{3w}{l^4}$$

Hence

$$|\Delta q| = \left|\frac{1}{l^3}\right| |\Delta w| + \left|-\frac{3w}{l^4}\right| |\Delta l|$$

$$= \frac{|\Delta w|}{l^3} + \frac{3w|\Delta l|}{l^4}$$

Using the data given we have

$$q = \frac{300}{(32)^3} = 9.2 \times 10^{-3}$$

and $$|\Delta q| \leq \frac{5}{(32)^3} + \frac{900.1}{(32)^4} = 1.0 \times 10^{-3}$$

Thus the maximum and minimum values of q are 10.2×10^{-3} and 8.2×10^{-3} respectively.

In this simple example we may check these extreme values by taking the errors in l and w in the most unfavourable ways. Thus l could be as low as 31 cm and w as high as 305 gm and this would lead to the maximum value of q =

$$q(\max) = \frac{305}{(31)^3} = 10.2 \times 10^{-3}$$

Similarly $l = 33$ and $w = 295$ gives the minimum value of q:

$$q(\min) = \frac{295}{(33)^3} = 8.2 \times 10^{-3}$$

These values are in agreement with the values found by adding and subtracting the estimated $|\Delta q|$ to $q = 9.2 \times 10^{-3}$.

The relative error in q is

$$\left|\frac{\Delta q}{q}\right| \leq \frac{l^3}{w}\left\{\frac{|\Delta w|}{l^3} + \frac{3w|\Delta l|}{l^4}\right\}$$

$$= \frac{|\Delta w|}{w} + \frac{3|\Delta l|}{l}$$

$$= \frac{5}{300} + \frac{3}{32}$$

$$= 1.1 \times 10^{-1} \quad (\text{or } 11\%)$$

Example 5.5.2

How accurate must l and w be if the maximum relative error in q should not exceed 1%?

We require

$$\frac{|\Delta q|}{q} \leq \frac{|\Delta w|}{w} + \frac{3|\Delta l|}{l} \leq 0.01$$

Clearly the values of $|\Delta w|$ and $|\Delta l|$ which satisfy this inequality are not unique. If, for example, we choose to make the terms $|\Delta w|/w$ and $3|\Delta l|/l$ contribute equally to the total error we must have

$$\frac{|\Delta w|}{w} \leq 0.005 \quad \text{and} \quad \frac{3|\Delta l|}{l} \leq 0.005$$

Inserting $w = 300$ and $l = 32$ in these expressions gives

$$|\Delta w| \leq 1.5 \text{ gm} \quad \text{and} \quad |\Delta l| \leq 0.053 \text{ cm}$$

If, on the other hand, we wish to make the relative errors in w and l the same, i.e. if we make $|\Delta w|/w = |\Delta l|/l = e$ (say), we must have

$$\frac{|\Delta q|}{q} \leq e + 3e = 4e \leq 0.01$$

giving $e \leq 0.0025$. This leads to

$$\frac{|\Delta w|}{w} \leq 0.0025 \text{ and } \frac{3|\Delta l|}{l} \leq 0.0025$$

or $$|\Delta w| \leq 0.75 \text{ gm and } |\Delta l| \leq 0.080 \text{ cm}$$

Both these solutions demand great accuracy in the measurement of length. Even if it were possible to measure w without error, l would need to be measured to the nearest mm to give q accurate to 1%. For in this case we would require

$$\frac{3|\Delta l|}{32} \leq 0.01 \quad \text{or} \quad |\Delta l| \leq 0.11 \text{ cm}$$

As a value of $|\Delta l| \leq 0.10$ cm is about the best that can be achieved when measuring a herring we see that

$$\frac{|\Delta w|}{w} \leq |\Delta q| - \frac{3|\Delta l|}{l} = 0.0100 - 0.0094 = 0.0006$$

or $$|\Delta w| \leq 0.18 \text{ gm}$$

That is, in order to estimate the condition factors of (large) herring to within 1%, the lengths must be accurate to the nearest mm and the weights to the nearest 0.2 gm.

As a further example of the use of the theory given in this section we give the derivation of the result quoted in section 3 for the maximum error of the sum of squares $\sum(w - \bar{w})^2$.

Here we have

$$u = \sum(w - \bar{w})^2 = \sum w^2 - (\sum w)^2/n$$

and $$\frac{\partial u}{\partial w_i} = 2w_i - 2(\sum w)/n \qquad (i = 1, 2, \ldots, n)$$

$$\therefore \quad \Delta u = \sum_{i=1}^{n} \left\{ \left| 2w_i - 2(\sum w)/n \right| . |\Delta w_i| \right\}$$

which is the required result.

5.6 Errors in the solution of simultaneous equations

Simultaneous linear equations arise frequently in mathematical and statistical calculations. In this section we use the preceding theory to estimate the errors in the solution of the two equations,

$$ax + by = c$$
$$dx + ey = f$$

in which the known quantities a, b, c, d, e and f are inaccurate.

The solutions of the above equations are

$$x = \frac{ce - bf}{ae - bd} = \frac{D'}{D} \quad \text{(say)}$$

$$y = \frac{af - cd}{ae - bd} = \frac{D''}{D} \quad \text{(say)}$$

The first-order partial derivatives of x and y are

$$\frac{\partial x}{\partial a} = -\frac{xe}{D} \qquad \frac{\partial y}{\partial a} = \frac{f}{D} - \frac{ye}{D}$$

$$\frac{\partial x}{\partial b} = -\frac{f}{D} + \frac{xd}{D} \qquad \frac{\partial y}{\partial b} = \frac{yd}{D}$$

$$\frac{\partial x}{\partial c} = \frac{e}{D} \qquad \frac{\partial y}{\partial c} = -\frac{d}{D}$$

$$\frac{\partial x}{\partial d} = \frac{xb}{D} \qquad \frac{\partial y}{\partial d} = -\frac{C}{D} + \frac{yb}{D}$$

$$\frac{\partial x}{\partial e} = \frac{C}{D} - \frac{xa}{D} \qquad \frac{\partial y}{\partial e} = -\frac{ya}{D}$$

$$\frac{\partial x}{\partial f} = -\frac{b}{D} \qquad \frac{\partial y}{\partial f} = \frac{a}{D}$$

To illustrate the use of these results let us take as a particular example the equations

$$5x + 3y = 11$$

$$2x + y = 4$$

the solutions of which are

$$x = \frac{D'}{D} = \frac{-1}{-1} = 1 \quad \text{and} \quad y = \frac{D''}{D} = \frac{-2}{-1} = 2$$

The first-order derivatives are

$$\left\{ \frac{\partial x}{\partial a}, \frac{\partial x}{\partial b}, \frac{\partial x}{\partial c}, \frac{\partial x}{\partial d}, \frac{\partial x}{\partial e}, \frac{\partial x}{\partial f} \right\} = \left\{ 1, 2, -1, -3, -6, 3 \right\}$$

$$\left\{ \frac{\partial y}{\partial a}, \frac{\partial y}{\partial b}, \frac{\partial y}{\partial c}, \frac{\partial y}{\partial d}, \frac{\partial y}{\partial e}, \frac{\partial y}{\partial f} \right\} = \left\{ -2, -4, 2, 5, 10, -5 \right\}$$

Let us assume that all the quantities a, b, ..., f have errors which are less than or equal to 0.01. Then the maximum absolute errors of $\underline{x}$ and $\underline{y}$ are

$$|\Delta x| \leq (1 + 2 + 1 + 3 + 6 + 3) \cdot 0.01 = 0.16$$

$$|\Delta y| \leq (2 + 4 + 2 + 5 + 10 + 5) \cdot 0.01 = 0.28$$

The signs of the first-order derivatives above indicate the combinations of the errors that lead to the most unfavourable situation and they also show that the equations which give maximum $\underline{x}$ give minimum $\underline{y}$ and vice versa. The equations that give maximum $\underline{x}$ and minimum of $\underline{y}$ are

$$5.01x + 3.01y = 10.99$$

$$1.99x + 0.99y = 4.01$$

with the solutions $\underline{x} = 1.16$ and $\underline{y} = 1.73$

The equations that give minimum $\underline{x}$ and maximum $\underline{y}$ are

$$4.99x + 2.99y = 11.01$$

$$2.01x + 1.01y = 3.99$$

with the solutions $\underline{x} = 0.84$ and $\underline{y} = 2.29$

We see that here also there is good agreement with the approximate estimates of error.

5.7 The statement of error

The maximum error of a function is used to state the reliability of the value of a function calculated from inaccurate values of the arguments. For the solutions of the simultaneous equations discussed in the previous section we write

$$x = 1.00 \pm 0.16$$

$$y = 2.00 \pm 0.28$$

For the condition factor evaluated in section 5.1 we write $q = 0.0092 \pm 0.0010$. Only a limited number of figures should be used and it is usual to take as the last figure in the error that after the first one in the result affected by both adding and subtracting the maximum error, just as has been done here.

We have discussed the problem of determining the maximum error of a function from a knowledge of the maximum errors of the arguments. The problem of using inaccurate data can be looked upon in a different way. As we seldom know the exact values of the arguments, the errors can have all values in intervals symmetrical about zero, and what we have done hitherto is to use the maximum errors combined in the most unfavourable way. It is clear that the maximum error of the function in question will be obtained only very seldom, and it would therefore be useful if it were possible to operate with some sort of most probable error.

This is possible and the only assumption required is that the errors of the arguments in question be randomly, symmetrically and independently distributed about zero.

To clarify the situation we use some statistical concepts, which will be dealt with properly in another chapter in this series. We shall work with variances and standard deviations, which will be denoted by $V(\cdot)$ and $\sigma(\cdot)$ respectively.

The variance of a function, f, can be expressed in terms of the variances of the arguments by means of formula (5.5.1) which gives

$$V(f) \sim \left\{\frac{\partial f}{\partial x}\right\}^2 \cdot V(x) + \left\{\frac{\partial f}{\partial y}\right\}^2 \cdot V(y) + \ldots.$$

The standard deviation measures dispersion about the true value in the following way:

The interval $(f - \sigma_f, f + \sigma_f)$ contains the true value with a probability of 68% or if the true value ϕ of f is known, the interval $(\phi - \sigma_f, \phi + \sigma_f)$ contains 68% of the f observations.

For intervals constructed using $2\sigma_f$ instead of σ_f we have an analogous statement with 95% probability instead of 68%.

We can in this way state the reliability of the calculated f by means of the standard deviation of f, and as the standard deviation and the maximum error operate, to a certain degree, in the same manner we write $f \pm \sigma_f$. A statement such as $f = 26.25 \pm 0.32$ has thus more than one sense and it is important always to state clearly what is meant.

For a sum

$$y = x_1 + x_2 + \ldots + x_n$$

the variance is

$$V(y) = V(x_1) + \ldots V(x_n)$$

and if all variances are equal

$$V(y) = nV(x)$$

For the standard deviation we find

$$\sigma_y = \sqrt{n}\,\sigma_x$$

For the maximum error of a sum we found earlier that it increases proportionally to n, and the factor $\sqrt{n}$ for the standard deviation can be taken as an effect of the cancelling of errors.

The theory of variances as used in calculations is to a large extent built up on the assumption that the errors are normally distributed, but even very crude approximations to normality in the errors of the argument give f variances that can be dealt with in the way just sketched. This is illustrated in the following example.

Example 5.7.1

Let us suppose that the errors of the coefficients in the equations of section 5.6 have the distributions

$$= \quad 0.01 \text{ with probability } 1/2$$

$$= -0.01 \quad - \qquad - \qquad 1/2$$

It is then clear that with six coefficients $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$ different sets of coefficients are possible and each of these have the probability 1 : 64. The solutions of the 64 sets of equations are:

x	y	x	y	x	y	x	y	x	y	x	y
0.96	2.06	1.04	1.94	0.92	2.12	0.96	2.09	1.13	1.77	0.91	2.15
0.94	2.10	1.02	1.98	0.90	2.16	1.04	1.91	1.12	1.80	1.02	1.96
0.92	2.15	1.09	1.83	0.87	2.21	1.02	1.96	1.10	1.84	1.00	2.00
0.90	2.19	1.07	1.87	0.86	2.25	1.00	2.00	0.90	2.16	0.98	2.04
0.98	2.02	1.06	1.90	0.96	2.06	0.98	2.04	0.88	2.20	0.96	2.08
0.96	2.06	1.04	1.94	0.94	2.10	1.07	1.87	0.85	2.25	1.04	1.93
0.94	2.11	1.14	1.77	0.91	2.16	1.05	1.91	0.84	2.29	1.02	1.96
0.92	2.14	1.12	1.81	0.89	2.20	1.00	2.00	1.02	1.98	1.04	1.94
1.02	1.96	1.10	1.84	0.98	2.02	0.98	2.04	1.10	1.83	1.06	1.90
1.00	2.00	1.08	1.88	0.96	2.06	1.08	1.86	1.08	1.87	1.04	1.94
0.98	2.04	1.16	1.73	0.93	2.11	1.06	1.90				

The variance of each of the coefficients can be shown to be 0.0001 and

$$V(x) = (1^2 + 2^2 + 1^2 + 3^2 + 6^2 + 3^2)\ 0.0001 = 0.0060$$

$$V(y) = (2^2 + 4^2 + 2^2 + 5^2 + 10^2 + 5^2)\ 0.0001 = 0.0174$$

Direct calculations give

Mean x = 0.998 V(x) = 0.0060

Mean y = 2.004 V(y) = 0.0175

in good agreement with the approximations.

The standard deviations are

$$\sigma_x = 0.077$$

$$\sigma_y = 0.132$$

The number of x values between (1.000 - 0.077) = 0.923 and (1.000 + 0.077) = 1.077 is 40 which is 62.5% of the 64. The number of x values between (1.000- 2 x 0.077) = 0.846 and (1.000 + 2 x 0.077) = 1.154 is 62 which is 96.9% of the 64. This is in good agreement with the figures stated for normal distributions, and the same is true for the y's. Using maximum errors then gives

$$x = 1.00 \pm 0.16$$

$$y = 2.00 \pm 0.28$$

as the solutions of the equations, whereas standard deviations give

$$x = 1.00 \pm 0.077$$

$$y = 2.00 \pm 0.132$$

The results are based on the same data and demonstrate the necessity of stating the exact meaning of the accuracies. The maximum error method is only used for rough statements whereas the standard deviation, which gives more information, is used in more refined calculations.

FAO FISHERIES TECHNICAL PAPER (MFS)

Documents in this group are provisional editions of volumes or sections of FAO MANUALS IN FISHERIES SCIENCE, usually in the languages of the original draft. They are given a limited distribution for comment by collaborators in this project, and for use of FAO headquarters and field staff and at training centers, courses and seminars. Revised versions are subsequently published by FAO in the official languages of the Organization.

FB/T26	Manual of Sampling Methods for Fisheries Biology	Oct. 1962
FIB/T40	Manual of Methods of Fish Population Analysis	Apr. 1964
FIb/T38	Tables of Yield Functions for Fishery Assessment	Jun. 1964
FIb/T26 Suppl. I	Manual of Sampling and statistical methods for fisheries biology. Part II - Statistical Methods - Chapter No. 5. Computations	Feb. 1965

A printed English version of " Manual of Field Methods in Fisheries Biology. Provisional Edition " was issued in 1960, as a revision of:

FB/58/T1	A Manual of Field Methods in Fisheries Biology. (Preliminary draft)	Sept. 1958

Drafts of " Manual of Laboratory Methods in Fisheries Biology " were issued as:

FB/58/T3	Part I - Physical and Chemical Methods in Meteorology and Oceanography	Jan. 1958
	Part II - Laboratory Methods on Plankton and Benthos	Feb. 1958

Revised versions of these have been combined with the Field Methods Manual to produce a " Manual of Methods in Fisheries Biology " the definitive text of which will be published in three language versions as FAO Man. Fisch. Sci., (1) (in preparation).

A provisional English edition of *FAO Man. Fish. Sci.*, (2) " Fisheries Science " was issued for FAO by CSIRO, Cronulla, Australia, in 1960.

24837/5.65/E/1/800